

Progress Toward a Forensic Voice Data Format Standard

James L. Wayman^{1*}, Joseph P. Campbell^{2†}, Pedro Torres-Carrasquillo^{2†}, Peter T. Higgins^{3*},
Alvin Martin^{4*}, Hirotaka Nakasone⁵, Craig Greenberg^{4*}, and Mark Przybocki^{4*}

162nd Meeting of the Acoustical Society of America
Forensic Acoustics Subcommittee
San Diego, CA
3 Nov 2011

¹ Office of Graduate Studies and Research, San José State University, San Jose, CA, USA

² Human Language Technology, MIT Lincoln Laboratory, Lexington, MA, USA

³ Higgins and Associates, International, Washington, DC, USA

⁴ Information Access Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

⁵ Digital Evidence Section, Operational Technology Division, Federal Bureau of Investigation, Quantico, VA, USA

*The work is sponsored by the Federal Bureau of Investigation under U.S. Department of Defense Contract J-FBI-07-164, and under U.S. Department of Commerce Contract A1101437. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

†This work is sponsored by the Federal Bureau of Investigation under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

The ANSI/NIST-ITL Standard

- Origins in 1980s motivated by needs of the FBI
- Now focused on law Enforcement, Military, Intelligence, Border Management and Homeland Security applications
- The *de facto* international standard for law enforcement exchange of biometric data
- Written very broadly, with specifics determined by exchange agreements within domains of interest
 - This impacts our approach to speech!

Locations of systems with the ANSI/NIST-ITL standard installed for biometric data transfer (known to NIST)



Blue: National and International System Use
Red: State / Provincial / Local System Use

Evolution of the ANSI/NIST-ITL Standard

- 1986, 1993, 1997 focused on fingerprint image and minutiae transmission
- 2000 version (Traditional format) still used in some applications: fingerprints & palm prints (images and minutiae), face images, scar-mark-tattoo images
- 2007 version (Traditional format) added iris and some new fingerprint minutiae fields
- 2008 version (XML format to meet legislative mandates) – same content as 2007 version

Today's ANSI/NIST-ITL 1-2011

- What's New :
 - modalities and data formats : DNA, Plantar (footprint), Iris compact formats, Images of additional body parts (besides face)
 - data: geo-positioning location; information assurance features; associated contextual images, audio or visual clips, and data
 - logs: data handling logs and original representation, audio or visual clips, DNA electropherograms and data (used to prepare the biometric samples)
- No provision for the exchange of voice data, but “Type 11 record” designation has been reserved
- No current record type suitable for adaption to voice

Speech is Different from Other Modes

- Time domain signal with changing characteristics
 - Background noise
 - Speaker behavior (pitch, vocal effort, emotional state ...)
 - Geographical location
- Naturally occurring in social context to exchange information
 - Multiple speakers
 - Semantic content
 - Code switching and social adaption
- Uncontrolled collection variables
 - Multiple devices
 - Transmission channel
 - Room acoustics

Getting Started on the Type 11 Record

- March 1-3, 2011 ANSI/NIST ITL-1/2 Workshop
 - Established a voice record development committee chaired by Dr. Alvin Martin (NIST)
- Task: define a *Transmission Standard* to exchange
 - voice biometric data
 - textual information fields relevant to processing and understanding the sample
- Type 11 will sit within a “transaction” that includes other records
 - Type 1 (transaction description)
 - Type 2 (data subject information)
- Transaction pertains to a single person, but may contain multiple voices
- Accommodate both segmented snips and diarizations?
- Best practices for controlled collections?
- Type 11 could contain the audio record, point to a Type 20 record within the transaction, or give external URL

Determining the Metadata

- How much voice in the sample?
- How, when and where was the recording made?
- How to present speech segment(s) inside an audio sample?
- Differentiate between permanent vs. temporary speaker attributes
- Differentiate between signal information and content information
- Quality metrics?

Type 11 Record Development

(way forward)

- Transition the committee proposal to the larger voice community for comment and revision
 - Is proposal useful over a broad range of use-cases?
- Utilize NIST ANSI/NIST-ITL expertise to formalize the Type 11 Record
- To be put forward as an amendment to ANSI/NIST-ITL 2011
- Proposal, comment, acceptance process will follow the standard ANSI balloting procedures